

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12Q 1/02, C12N 15/15, 1/21</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 93/01305</b> <b>(43) International Publication Date:</b> 21 January 1993 (21.01.93)
<b>(21) International Application Number:</b> PCT/US92/05745 <b>(22) International Filing Date:</b> 9 July 1992 (09.07.92)  <b>(30) Priority data:</b> 727,837 9 July 1991 (09.07.91) US  <b>(71)(72) Applicant and Inventor:</b> BALINT, Robert [US/US]; 1823 Rigeland Circle, Danville, CA 94526 (US).  <b>(74) Agents:</b> MURASHIGE, Kate, H. et al.; Morrison & Foerster, 755 Page Mill Road, Palo Alto, CA 94304 (US).  <b>(81) Designated States:</b> AU, CA, JP, KR, RU, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).		<b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> A METHOD FOR THE IDENTIFICATION OF PROTEASE INHIBITORS  <b>(57) Abstract</b>  Methods for detecting protease inhibitors are disclosed. Also described are DNA constructs and host cells transformed with these constructs for use in the subject methods. The methods utilize a host cell which exhibits a negative phenotype dependent on the activity of a given protease. Thus, inhibition of the protease confers a selectable phenotype on the cell. The negative phenotype can be conferred by either protease-mediated activation or inactivation of a protein conferring a selectable phenotype. The inhibitor is detected by transforming host cells expressing the genes for the selectable phenotype and a given protease with random peptide sequences. Inhibitors so identified can be used either directly or indirectly in the treatment of protease-dependent disorders.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MI	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NI	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

5     A METHOD FOR THE IDENTIFICATION OF PROTEASE INHIBITORSTechnical Field

          The instant invention relates generally to the identification of protease inhibitors. More  
10     particularly, the invention relates to methods of identifying viral protease inhibitors which can in turn be used to treat or prevent viral infection.

Background

15           Proteases are enzymes that cleave peptide bonds, thereby altering proteins. Besides degrading proteins, these enzymes play a regulatory role in a variety of physiological processes. Proteases fall into four general classes: serine, cysteine, aspartic acid,  
20     and metalloproteases. These classes are distinguished primarily by mechanism (Dunn, B.M., in Proteolytic Enzymes, R.J. Geynon and J.S. Bond, eds., IRL Press, Oxford, 1989, pp. 57-82). Serine and cysteine proteases have almost identical two-step mechanisms with an acyl-  
25     enzyme intermediate. Together they comprise the majority of the known proteases. Aspartic and metallo-proteases catalyze direct hydrolysis of the peptide bond.

          Many procaryotic and eucaryotic proteins are synthesized as larger biologically inactive precursors  
30     which become activated only when acted upon by endoproteases (proteinases). These enzymes typically recognize specific domains, usually less than ten amino acids in length including the scissile bond, in exposed  
35     loops of generally loose secondary structure (Keil, B.,

-2-

in Methods in Protein Sequence Analysis, M. Elzinga, ed., Humana Press, Clifton, N.J., 1982, pp. 291-304).

Maturation proteases are responsible for both intracellular and extracellular cleavage of protein precursors and many of the proteolytically processed proteins in turn play key roles in physiological abnormalities which give rise to disease states (Andrews, P.C., et al., Experientia (1987) 43:784-789; Reich, E., et al., Cold Spring Harbor Symposium: Proteases and Biological Control, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1975). Proteins that undergo intracellular proteolytic maturation include secreted proteins, lysosomal enzymes, mitochondrial proteins and membrane proteins. These proteins are highly diverse in function, having endocrine, neurological, and immune functions, as well as acting as growth factors and antibiotics. Secreted proteins that undergo extracellular proteolytic processing include the plasma zymogens involved in blood clotting and the immune complement system.

Maturation proteases which are indirectly involved in human disease are generally distinguished by their high degree of substrate specificity. However, a host of digestive proteases of lesser specificity are also known and are more directly involved in diseases such as chronic inflammation and tumor metastasis. These enzymes include elastases, collagenases, mast cell proteases, and extracellular matrix-degrading metalloproteases, among others.

Proteases also play key roles in many infectious diseases. An obligatory step in the replication of many pathogenic plant and animal viruses involves virus-determined proteolytic processing of the primary viral gene products (Hellen, C.U.T., et al.,

**SUBSTITUTE SHEET**

-3-

Biochemistry (1989) 28:9881-9890). Plant viruses which encode proteases for this purpose include the potyviruses, comoviruses, nepoviruses, sobemoviruses, and luteoviruses. These viruses cause economically important diseases in all major monocot and dicot families. Similarly equipped animal viruses include the picornaviruses, retroviruses, alphaviruses, flaviviruses, pestiviruses, coronaviruses, and adenoviruses. Diseases caused by these viruses include foot-and-mouth disease, AIDS, the common cold, hepatitis and polio.

For example, Zucchini Yellow Mosaic Virus (ZYMV), a potyvirus, expresses its genome as a single 350 kDa polyprotein which is cleaved into at least seven mature gene products by three distinct proteolytic activities. Two of the proteases are virus-encoded (Dougherty, W.G., and J.C. Carrington, Ann. Rev. Phytopathol. (1988) 26:123-143; Carrington, J.C., et al., EMBO J. (1990) 9:1347-1353), including the potyviral 49 kDa protease. This protease is responsible for at least five of the seven cleavages. This enzyme is a trypsin-like cysteine protease which is structurally and mechanistically representative of the largest class of viral proteases, including those of the animal picornaviruses (Dougherty, W.G., et al., Virology (1989) 172:302-310; Bazan, J.F., and R.J. Fletterick, Proc. Natl. Acad. Sci. USA (1988) 85:7872-7876). This enzyme is highly specific and appears to recognize a region comprised of about seven amino acids surrounding the scissile bond (Dougherty, W.G., and T.D. Parks, Virology (1989) 172:145-155). Of the five sites cleaved by this enzyme, the two flanking the protease appear to be cleaved intramolecularly, while the remaining three appear to be cleaved intermolecularly (Garcia, J.A., et al., J. Gen. Virol. (1990) 71:2773-2779). Of the latter

SUBSTITUTE SHEET

-4-

three, the site between the N1b protein and the coat protein appears to be the most active.

#### Disclosure of the Invention

5           The invention herein is based on the discovery of a unique method for detecting peptide protease inhibitors. These inhibitors can be used directly or indirectly in the treatment of protease-dependent diseases. Alternatively, the inhibitors so identified  
10 can be utilized as structural models for the rational design of peptide-mimetics.

Accordingly, in one embodiment, the subject invention is directed to a method for detecting a protease inhibitor which comprises:

- 15           (a) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on said host cells dependent on the activity of said  
20 protease;
- (b) providing a pool of nucleic acid constructs wherein at least one of the constructs in the pool comprises a nucleic acid sequence encoding an inhibitor of the protease;
- 25           (c) transforming the host cells of (a) with the nucleic acid constructs of (b); and
- (d) growing the transformed host cells of (c) under conditions that distinguish cells with the selectable phenotype, thereby detecting the presence of  
30 the protease inhibitor.

In another embodiment, the subject invention is directed to a DNA construct comprising:

- 35           (a) a first DNA coding sequence for a protein capable of conferring a selectable phenotype on a host

-5-

cell transformed therewith, the selectable phenotype dependent on the activity of a protease; and

(b) control sequences that are operably linked to the first and second coding sequences whereby the coding sequences can be transcribed and translated in a host cell, and at least one of the control sequences is heterologous to at least one of the coding sequences.

In an alternate embodiment, the DNA construct further includes a second DNA coding sequence for the protease of interest.

In yet another embodiment, the subject invention is directed to host cells stably transformed with these DNA constructs.

These and other embodiments of the subject invention will readily occur to those of ordinary skill in the art in view of the disclosure herein..

#### Brief Description of the Figures

Figure 1 depicts Protease Inhibitor Selection System I, as applied to ZYMV protease.

Figure 2 depicts representative examples of Protease Inhibitor Selection System II.

Figure 3 shows the strategy of cDNA synthesis from ZYMV and cloning methods.

Figure 4 shows the nucleotide sequence of the ZYMV genome (SEQ ID NO:1).

Figure 5A shows the organization of the primary translation products of pZPro6, pZPro7 and placZa-CP. Figure 5B depicts the results of immunoblot analysis of SDS/PAGE separated proteins from *E. coli* cells harboring these plasmids.

Figure 6 depicts the derivation of the pZPro7, pZPro9, pZPro10, pZPro11 and pZPro12 constructs and the organization of the primary translation products. The

-6-

open boxes denote ZYMV 49 kDa protease (Pro) cleavage sites. Strep<sup>R</sup> = streptomycin-resistant. Amp<sup>R</sup> = ampicillin-resistant, i.e., transformed. CfU = colony-forming units. NT = not tested.

5

#### Detailed Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, virology, recombinant DNA technology, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989); Maniatis, Fritsch & Sambrook, Molecular Cloning: A Laboratory Manual (1982); DNA Cloning, Vols. I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds 1984); Animal Cell Culture (R.K. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the series, Methods In Enzymology (S. Colowick and N. Kaplan eds., Academic Press, Inc.); and Handbook of Experimental Immunology, Vols. I-IV (D.M. Weir and C.C. Blackwell eds., 1986, Blackwell Scientific Publications).

20  
25

All patents, patent applications, and publications mentioned herein, whether supra or infra, are hereby incorporated by reference in their entirety.

#### 30 A. Definitions

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

35



-7-

By "protease" is meant an enzyme that cleaves a peptide bond. The term includes both endopeptidases (also called proteinases) which are proteases that hydrolyze internal peptide bonds, and exopeptidases, which are proteases that cleave either N- or C-terminal peptide bonds. Some proteases are highly specific, cleaving only between two particular amino acids within a particular protein. Other proteases are less specific, cleaving between more than one amino acid pair and/or cleaving between an amino acid pair in more than one location in the same and/or different proteins. Exemplary proteases include maturation proteases responsible for both intracellular and extracellular cleavage of protein precursors, such as secreted proteins, lysosomal enzymes, mitochondrial proteins, membrane proteins, plasma zymogens, digestive enzymes, elastases, collagenases, mast cell proteases, extracellular matrix-degrading metalloproteinases; plant viral proteases such as proteases from potyviruses, comoviruses, nepoviruses, sobemoviruses, and luteoviruses; and animal viral proteases such as proteases from picornaviruses, retroviruses, alphaviruses, flaviviruses, pestiviruses, coronaviruses, and adenoviruses.

By "protease inhibitor" is meant a molecule capable of altering the activity of a protease such that the protease is unable to completely hydrolyze a peptide bond for which it is specific. Protease inhibitors can be peptides composed solely of genetically encodable amino acids. "Protease inhibitor" also encompasses synthetic peptide derivatives such as peptide aldehydes and ketones, peptide boronic acids, peptide chloromethyl ketones, azapeptides, peptide hydroxamic acids, and peptide thiols. "Protease inhibitor" also encompasses

**SUBSTITUTE SHEET**

-8-

synthetic nonpeptide compounds such as diisopropylphosphofluoridate, sulfonylfluorides, phosphoramidon, and halomethylcoumarins. For a detailed discussion of protease inhibitors, see Proteinase Inhibitors, A.J. Barrett and G. Salvesen, eds. (Elsevier, Amsterdam, 1986).

The terms "peptide" and "protein" are used in their broadest sense, i.e., any polymer of genetically encodable amino acids (dipeptide or greater) linked through peptide bonds. Thus, the terms include oligopeptides, polypeptides, protein fragments, muteins, fusion proteins and the like.

A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous nucleotide sequence. As described more fully below, host cells for use in the present invention may be either procaryote or eucaryote, depending on the specific protease in question and the selection system desired. In general, bacterial cells (either gram-negative or gram-positive) are the hosts of choice when the protease and its dependent phenotype can be expressed in active form in these cells. Eucaryotic cells can be used, however, in cases where either the protease or its dependent phenotype can be adequately expressed only in such cells, such as cases in which certain types of transport, metabolism, or post-translational modification are required. Eucaryotic cells can also be used to select inhibitors of other types of biological activities which can be expressed only in such cells, such as animal virus replication. One skilled in the art can readily determine an appropriate host cell for use in the present invention.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an

**SUBSTITUTE SHEET**

-9-

autonomous unit of DNA replication; i.e., capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (bases adenine, guanine, thymine, or cytosine) in a double-stranded helix, both relaxed and supercoiled. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence homologous to the mRNA).

A DNA "coding sequence" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, procaryotic sequences, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

**SUBSTITUTE SHEET**

-10-

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at the 3' terminus by the translation start codon (ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eucaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Procaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell.

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into mRNA, which is then translated into a chimeric polypeptide encoded by the two coding sequences. The coding sequences need not be contiguous to one another so long as the transcribed sequence is ultimately processed to produce the desired chimeric protein.

A control sequence "directs the transcription" of a coding sequence in a cell when RNA polymerase will

**SUBSTITUTE SHEET**

-11-

bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A cell has been "transformed" by an exogenous nucleotide sequence when the sequence has been introduced inside the cell membrane. An exogenous nucleotide sequence may or may not be integrated (covalently linked) to chromosomal nucleic acid making up the genome of the cell. In procaryotes and yeasts, for example, the exogenous nucleotide sequence may be maintained on an episomal element, such as a plasmid. With respect to most other eucaryotic cells, a stably transformed cell is one in which the exogenous nucleotide sequence has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eucaryotic cell to establish cell lines or clones comprised of a population of daughter cell containing the exogenous sequence.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature. Thus, when the heterologous region encodes a bacterial gene, the gene will usually be flanked by DNA that does not flank the bacterial gene in the genome of the source bacteria. Another example of the heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Allelic variation or

**SUBSTITUTE SHEET**

-12-

naturally occurring mutational events do not give rise to a heterologous region of DNA, as used herein.

The term "treatment" as used herein refers to either (i) the prevention of infection or reinfection  
5 (prophylaxis), or (ii) the reduction or elimination of symptoms of the disease of interest (therapy).

#### B. General Methods

Described herein is a system which can be used  
10 to select effective inhibitors of proteases from large pools of random peptide sequences. The method utilizes a known protease which can be obtained through standard techniques, i.e. direct isolation, synthesis or  
15 recombinant technology. The nucleotide sequence of the protease can be determined and used to transform a host cell. The host cell is also transformed with a nucleotide sequence encoding a protein that confers a negative phenotype on the cell, such as sensitivity to a given antibiotic, or inability to grow on a given carbon  
20 source, which is dependent on the activity of the cloned protease. Genes for the protease and the protein conferring the dependent phenotype are contained on one or more constructs which have been introduced into the host cell. Thus, inhibition of the protease confers a  
25 selectable phenotype on the cell (e.g., antibiotic resistance, or the ability to grow on a given carbon source). Once identified, the particular inhibitor can be isolated, sequenced and further used as described below.

30 The negative phenotype may be expressed by either of two general mechanisms. In the first, a gene conferring a dominant negative phenotype is expressed as an inactive precursor protein which is activated by  
35 protease-mediated cleavage at a site which has been

**SUBSTITUTE SHEET**

-13-

engineered to resemble a natural substrate of the protease. In the second mechanism, a gene conferring a selectable phenotype is inactivated by protease-mediated cleavage at a similarly engineered site.

5           The above-described host cells can then be used to detect effective inhibitors of the protease from large pools of random peptides encoded on another plasmid. Cells transformed with variants of this plasmid which encode effective inhibitors are identified by selection  
10   for the appropriate phenotype. This additional plasmid contains a gene which encodes a "carrier" protein in which all or part of an exposed domain has been randomized with respect to its amino acid sequence. Typically, the randomized domain may range from four to  
15   fifteen amino acids in length. The length of the randomized amino acid sequence will depend on the specific application of the inhibitor and can be readily determined by one skilled in the art. For example, peptides intended for use for the design of peptide  
20   mimetics will tend to have shorter sequences than peptides for use in peptide or gene therapy.

Part or all of this sequence is randomized with respect to the twenty genetically-encodable amino acids. Thus, a fully randomized set of heptapeptide sequences  
25   would contain more than  $10^9$  different peptides.

Such random sequence "libraries" can be constructed by replacing the sequence encoding the exposed domain in the "carrier" protein gene with a set of synthetic oligodeoxynucleotides of random sequence. A  
30   natural substrate of the protease in question can be conveniently used for the "carrier" protein. Alternatively, one of the many well-characterized natural protease inhibitors may be used (Proteinase Inhibitors,  
35   A.J. Barrett and G. Salvesen, eds. (Elsevier, Amsterdam,

**SUBSTITUTE SHEET**

-14-

1986), section C), in which the amino acid sequence of the native binding domain has been randomized. Structural constraints placed on the randomized sequence by the flanking domains of the "carrier" may be minimized by flanking the randomized sequence with short "spacers" of polyproline or polyglycine which are highly flexible (Creighton, T.E., Proteins: Structures and Molecular Properties, W.H. Freeman, New York, 1984, ch. 5).

Some of these "random" peptides will, by chance, have structures which are capable of binding tightly to the active site of the protease, thereby preventing it from either activating or inactivating the negative phenotype-conferring protein, depending on the mechanism employed. This, in turn, will confer the selectable phenotype on the host cell when transformed with these constructs. The structures of effective inhibitors can then be determined by sequencing the appropriate regions of constructs recovered from such phenotype-selected cells.

A representative example of the first mechanism described above, i.e., wherein the activation of a negative phenotype-conferring protein is inhibited (hereinafter referred to as Protease Inhibitor Selection System I) as applied to the ZYMV protease is illustrated in Figure 1. A portion of the ZYMV polyprotein is depicted which includes the protease, replicase (NIb), and coat protein (see U.S. Patent Application Serial No. 07/560,130). The arrows indicate substrate sites at which the protease cleaves the polyprotein. In this selection system, either the replicase or coat protein is replaced with the coding sequence for the protein conferring the negative phenotype. An example of the latter includes E. coli ribosomal protein S12, which confers sensitivity to streptomycin on streptomycin-

SUBSTITUTE SHEET



-15-

resistant hosts such as E. coli strain MC1009 (Post, L.E., and M. Nomura, J. Biol. Chem. (1980) 255:4660-4666). A transcription repressor protein may also be used such as the lactose, tryptophan, or phage lambda repressors (Lewin, B., Genes IV, Cell Press, Cambridge, 1990, pp. 240-264). These act by repressing expression of antibiotic resistance genes in hosts in which these genes are transcribed from repressible promoters. In either case the negative phenotype is only displayed when the negative phenotype protein is being actively cleaved out of the polyprotein by the protease.

For other proteases it may be convenient to express the protease and the negative phenotype precursor from separate transcription units. The only requirements are that the negative phenotype protein be linked to an extraneous domain by a peptide sequence which is a natural substrate for the protease in question, and that this precursor be inactive until cleaved by the protease.

Figure 2 illustrates the second mechanism wherein the negative phenotype is conferred by protease-mediated inactivation of a protein conferring a selectable phenotype (referred to herein as Protease Inhibitor Selection System II). Examples of such proteins include secreted or membrane proteins which confer resistance to the antibiotics ampicillin, tetracycline, or kanamycin (Methods in Enzymology, vol. 43, Academic Press, New York, 1975), or which confer the ability to utilize carbon sources such as lactose or maltose (Bieker, K.L., and T.J. Silhavy, Trends in Genetics (1990) 6:329-334). These proteins are normally expressed as precursors in which an amino-terminal signal sequence directs transport of the protein across the cell membrane or insertion of the protein into the cell membrane, after which the signal sequence is

**SUBSTITUTE SHEET**

-16-

proteolytically removed. In these constructs, the protease substrate peptide sequence is inserted between the signal sequence and the mature protein such that cleavage by the protease renders the protein incapable of membrane transport or insertion and thereby inactive. Alternatively, the protease substrate sequence may be inserted into a surface domain of the mature protein such that cleavage by the protease renders the protein inactive.

A special case of this selection system occurs with proteases which are toxic when expressed in E. coli by virtue of their fortuitous inactivation of one or more host proteins which are required for growth (Baum, E.Z., et al., Proc. Natl. Acad. Sci. USA (1990) 87:5573-5577).

In such cases, the inactivated host protein(s) confer the selectable phenotype in the presence of inhibitors of the protease.

The random peptide inhibitor gene library may be delivered to the selector cells by any of several methods, the choice of which will depend to some extent on the size of the library. One skilled in the art can readily determine an acceptable technique to use with a given library. For example, chemical transformation with purified plasmid (Sambrook, J., et al., Molecular Cloning, Cold Spring Harbor Laboratory, 1989, pp. 1.76-1.84) can be used for libraries of up to  $10^8$ - $10^9$  members, depending on the efficiency. Such a library can accommodate a complete set of fully randomized pentapeptides. High voltage electroporation with purified plasmid (Dower, W.J., et al., Nucleic Acids Res. (1988) 16:6127-6145) is useful for libraries of  $10^{10}$ - $10^{11}$  members, nearly sufficient to accommodate a complete set of fully randomized heptapeptides. For larger libraries,

**SUBSTITUTE SHEET**

-17-

bacteriophage-derived vectors can be used for delivery by transduction.

For example, a plasmid vector can be converted to a cosmid vector (Sambrook, J., et al., Molecular Cloning, Cold Spring Harbor Laboratory, 1989, ch. 3) simply by insertion of a cos site and an appropriate length of "stuffer" DNA. Concatenate ligation of the library to such a vector can be followed by efficient packaging into phage  $\lambda$  pseudovirions using commercially available preparations. Efficient, large-scale transductions of the packaged cosmids into selector cells can then be accomplished by established methods. In a further refinement, concatemers of the inhibitor gene library can be used instead of "stuffer" DNA in the cosmid to achieve the necessary size for packaging. This reduces, by an order of magnitude, the number of transformants that need to be screened to cover the entire library.

The stringency of selection by these systems can be adjusted in a variety of ways. A number of transcriptional promoters and enhancers of varying strengths are available (Sambrook, J., et al., Molecular Cloning, Cold Spring Harbor Laboratory, 1989, ch. 17), which can be used with the protease, negative phenotype precursor, and inhibitor genes to raise or lower the inhibitor strength required for selection. Inducible promoters can be used, such that their strengths may be titrated by adjusting the amount of inducer in the growth medium. For example, by having the inhibitor expressed from an inducible promoter, the potency threshold of a pool of selected inhibitors is raised, and the size of the pool reduced, simply by reducing the amount of inducer present during selection. In addition, such inhibitor inducibility can be used to counterselect

**SUBSTITUTE SHEET**

-18-

stable false positives, such as revertants that have mutated the protease gene, simply by replica plating onto selective medium in the absence of inducer. Only the revertants are able to grow.

5           Once detected, the protease inhibitor can be isolated and chemically characterized, using known techniques. These systems can be used to generate inhibitor peptides for any protease which can be expressed in active form in a suitable host and for which  
10   substrate cleavage site sequences are known. In addition to the proteases of many important plant and animal viral pathogens, inhibitors of the proteases of other types of microbial pathogens as well as cellular proteases which have been implicated in such disorders as rheumatoid  
15   arthritis, Alzheimer's disease, and tumor metastasis, can also be identified.

#### C. Use and Administration

          The instant invention can be used to identify  
20   protease inhibitors which in turn are useful in the treatment of protease-dependent diseases in both plants and animals. The inhibitors can be used directly in peptide therapy or can be encoded in a gene and used in gene therapy. The identified inhibitors can also serve  
25   as structural models for the rational design of peptide-mimetics, that is, synthetic compounds that mimic the protease-binding action of the identified protease inhibitors to bring reactive groups into contact with the protease active site. (See, e.g., Demuth, H.-U., J.  
30   Enzyme Inhibition (1990) 3:249-278). The present invention also has a more general application in the construction and use of in vivo systems for the selection of bioactive peptides from peptide libraries. For  
35   example, systems may be designed for the selection of

-19-

peptide inhibitors of hydrolytic enzymes other than proteases. A number of such enzymes are known which can hydrolyze natural or artificial substrates to produce one or more compounds which are toxic to E. coli (for  
5 example, see Hydrolytic Enzymes, A. Neuberger and K. Brocklehurst, eds. Elsevier, Amsterdam, 1987). The expression of such an enzyme in an appropriate host cell allows the selection of peptide inhibitors of the enzyme based on their ability to confer viability on the cells  
10 in the presence of toxigenic substrates.

In general, any phenotype of cultured procaryotic or eucaryotic cells which can be altered by the endogenous expression of appropriate peptides such that cells expressing such peptides can be readily  
15 distinguished and isolated from cells which either do not express such peptides or which express peptides which do not alter the phenotype, may provide the basis for establishing an in vivo system for the selection of bioactive peptides from peptide libraries. Among the  
20 most tractable medically important phenotypes will be those manifesting susceptibility to microbial pathogenicity.

For example, the endogenous expression of a random peptide library as an exposed domain of a suitable  
25 stable "carrier" protein in a population of cultured mammalian cells of sufficient size to ensure that all or most members of the library are represented in the population may be used to select peptides which interfere with the ability of microbial pathogens such as viruses  
30 or bacteria or their toxins to inhibit cell growth. When such cell populations are challenged by such pathogens or toxins, only those cells expressing inhibitory peptides will grow, allowing the active peptides to be identified  
35

**SUBSTITUTE SHEET**

-20-

by established methods. Such peptides can in turn be used for the development of effective therapies.

For the treatment of plant pathogenesis, the identified inhibitors can be used to create transgenic plants. One commonly used method of gene transfer in plants involves the insertion of the gene of interest into the T-DNA region of a Ti or Ri plasmid derived from A. tumefaciens or A. rhizogenes, respectively. Many control sequences are known which when coupled to a heterologous coding sequence and transformed into a host organism show fidelity in gene expression with respect to tissue/organ specificity of the original coding sequence. See, e.g., Benfey, P.N., and Chua, N.H., Science (1989) 244:174-181. Suitable control sequences for use in these plasmids include promoters for constitutive leaf-specific expression of the desired gene in the various target plants. Other useful control sequences include a promoter and terminator from the nopaline synthase gene (NOS). The NOS promoter and terminator are present in the plasmid pARC2, available from the American Type Culture Collection and designated ATCC 67238. If such a system is used, the virulence (vir) gene from either the Ti or Ri plasmid must also be present, either along with the T-DNA portion, or via a binary system where the vir gene is present on a separate vector. Such systems, vectors for use therein, and methods of transforming plant cells are described in U.S. Patent No. 4,658,082, and Simpson, R.B., et al., Plant Mol. Biol. (1986) 6:403-415, incorporated herein by reference in their entirety.

Once constructed, these plasmids can be placed into A. rhizogenes or A. tumefaciens and these vectors used to transform cells of plant species which are ordinarily susceptible to the particular plant pathogen. The selection of either A. tumefaciens or A. rhizogenes

**SUBSTITUTE SHEET**

-21-

will depend on the plant being transformed thereby. In general, A. tumefaciens is the preferred organism for transformation. Most dicotyledons, some gymnosperms, and a few monocotyledons (e.g., certain members of the Liliales and Arales) are susceptible to infection with A. tumefaciens. A. rhizogenes also has a wide host range, embracing most dicots and some gymnosperms, which includes members of the Leguminosae, Compositae and Chenopodiaceae. Alternative techniques which have proven to be effective in genetically transforming plants include particle bombardment and electroporation. See, e.g., Rhodes, C.A., et al., Science (1988) 240:204-207; Shigekawa, K., and Dower, W.J., BioTechniques (1988) 6:742-751; Sanford, J.C., et al., Particulate Science and Technology (1987) 5:27-37; and McCabe, D.B., BioTechnology (1988) 6:923-926.

Once transformed, these cells can be used to regenerate transgenic plants. For example, whole plants can be infected with these vectors by wounding the plant and then introducing the vector into the wound site. Any part of the plant can be wounded, including leaves, stems and roots. Alternatively, plant tissue, in the form of an explant, such as cotyledonary tissue or leaf disks, can be inoculated with these vectors and cultured under conditions which promote plant regeneration. Roots or shoots transformed by inoculation of plant tissue with A. rhizogenes or A. tumefaciens, containing the desired gene, can be used as a source of plant tissue to regenerate transgenic plants, either via somatic embryogenesis or organogenesis. Examples of such methods for regenerating plant tissue are disclosed in Shahin, E.A., Theor. Appl. Genet. (1985) 69:235-240; U.S. Patent No. 4,658,082; and Simpson et al., supra.

**SUBSTITUTE SHEET**

-22-

The inhibitors identified by the present method can also be used in gene therapy. For example, HIV-specific protease inhibitor genes, in which a natural mammalian protease inhibitor serves as carrier for the HIV protease inhibitor domain, can be used in anti-AIDS gene therapy. Lymphocytes or bone marrow cells from the patient can be transformed with the protease inhibitor gene in vitro and returned to the patient, where they establish an HIV-resistant subpopulation of lymphocytes which can gradually restore cell-mediated immune function as the patient's untransformed lymphocytes are depleted by the virus.

Similarly, proteases active in blood, lymph, or cerebro-spinal fluid which are essential components of disorders such as chronic inflammations, metastatic cancers, and certain viral infections, may be targeted by protease inhibitor gene therapy, in which the inhibitors are secreted by transgenic lymphocytes or other transgenic cell implants.

For therapeutic use in animals, the inhibitors identified by the present method can be altered by established methods to improve their pharmacokinetic properties. For example, the inhibitors may be administered linked to a carrier. For example, a fragment may be conjugated with a macromolecular carrier. Suitable carriers are typically large, slowly metabolized macromolecules such as: proteins; polysaccharides, such as sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, and other proteins well known to those skilled in the art.

SUBSTITUTE SHEET



-23-

The protein substrates may be used in their native form or their functional group content may be modified by, for example, succinylation of lysine residues or reaction with Cys-thiolactone. A sulfhydryl group may also be incorporated into the carrier (or inhibitor) by, for example, reaction of amino functions with 2-iminothiolane or the N-hydroxysuccinimide ester of 3-(4-dithiopyridyl) propionate. Suitable carriers may also be modified to incorporate spacer arms (such as hexamethylene diamine or other bifunctional molecules of similar size) for attachment of peptides. Methods of coupling peptides to proteins or cells are known to those of skill in the art.

It is also possible to administer the inhibitors identified using the instant method alone, or mixed with a pharmaceutically acceptable vehicle or excipient. Typically, the compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation may also be emulsified or the active ingredient encapsulated in liposome vehicles. The active immunogenic ingredient is often mixed with vehicles containing excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable vehicles are, for example, water, saline, dextrose, glycerol, ethanol, or the like, and combinations thereof. In addition, if desired, the vehicle may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, or pH buffering agents. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in the art. See, e.g., Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pennsylvania, 15th edition, 1975. The

**SUBSTITUTE SHEET**

-24-

composition or formulation to be administered will, in any event, contain a quantity of the inhibitor adequate to achieve the desired effect in the individual being treated.

5 Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, aerosol, intranasal, oral formulations, and sustained release formulations. For suppositories, the vehicle composition will include traditional binders  
10 and carriers, such as, polyalkylene glycols, or triglycerides. Such suppositories may be formed from mixtures containing the active ingredient in the range of about 0.5% to about 10% (w/w), preferably about 1% to about 2%. Oral vehicles include such normally employed  
15 excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium, stearate, sodium saccharin cellulose, magnesium carbonate, and the like. These oral compositions may be taken in the form of solutions, suspensions, tablets, pills, capsules,  
20 sustained release formulations, or powders, and contain from about 10% to about 95% of the active ingredient, preferably about 25% to about 70%.

Intranasal formulations will usually include vehicles that neither cause irritation to the nasal  
25 mucosa nor significantly disturb ciliary function. Diluents such as water, aqueous saline or other known substances can be employed with the subject invention. The nasal formulations may also contain preservatives such as, but not limited to, chlorobutanol and  
30 benzalkonium chloride. A surfactant may be present to enhance absorption of the subject proteins by the nasal mucosa.

Controlled or sustained release formulations  
35 are made by incorporating the inhibitor into carriers or

**SUBSTITUTE SHEET**

-25-

vehicles such as liposomes, nonresorbable impermeable polymers such as ethylenevinyl acetate copolymers and Hytrel® copolymers, swellable polymers such as hydrogels, or resorbable polymers such as collagen and certain  
5 polyacids or polyesters such as those used to make resorbable sutures. The inhibitors can also be delivered using implanted mini-pumps, well known in the art.

Furthermore, the inhibitors (or complexes thereof) may be formulated into pharmaceutical compositions in either neutral or salt forms. Pharmaceutically  
10 acceptable salts include the acid addition salts (formed with the free amino groups of the active polypeptides) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such  
15 organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine,  
20 trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

To treat an animal subject, the inhibitor of interest is administered parenterally, usually by intramuscular injection in an appropriate vehicle. Other  
25 modes of administration, however, such as subcutaneous, intravenous injection and intranasal delivery, are also acceptable. Injectable formulations will contain an effective amount of the active ingredient in a vehicle, the exact amount being readily determined by one skilled in  
30 the art. The active ingredient may typically range from about 1% to about 95% (w/w) of the composition, or even higher or lower if appropriate. The quantity to be administered depends on the animal to be treated and the  
35 particular inhibitor used. Effective dosages can be

**SUBSTITUTE SHEET**

-26-

readily established by one of ordinary skill in the art through routine trials establishing dose response curves. The subject is treated by administration of the particular inhibitor, in at least one dose. Moreover,  
5 the subject may be administered as many doses as is required to effectively treat the individual.

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not  
10 intended to limit the scope of the present invention in any way.

#### EXAMPLES

15

##### Example 1

The genetic expression of active ZYMV 49 kDa protease in *E. coli*.

This example describes the construction and expression in *E. coli* of a gene which encodes a portion  
20 of the ZYMV polyprotein. The primary translation product of this gene is a 140 kDa protein which includes the 49 kDa protease and flanking cleavage sites, a portion of the nuclear inclusion 'b' protein (NIb, also referred to as the replicase), including the NIb/coat protein  
25 cleavage site, followed by the coat protein (CP). Evidence is presented showing that the expression of this gene in *E. coli* leads to an accumulation of mature CP as a result of efficient cleavage at the NIb/CP cleavage site by the 49 kDa protease.

30

##### cDNA Cloning and Sequencing of the ZYMV Genome

A California isolate of ZYMV was obtained from Professor J.A. Dodds of the University of California at  
35 Riverside. The virus was propagated by mechanical

-27-

inoculation of the cotyledons of ten-day-old Cucurbita pepo cv. early straightneck seedlings. Systemically infected leaves were harvested 3-8 weeks after inoculation and virus was purified therefrom essentially as described by Lisa, V., et al., Phytopathol (1981) 71:667-672. The virus was quantified by absorbance at 260 nm using an extinction coefficient of 2.8 A<sub>260</sub>/mg/ml.

5  
10  
15  
Viral genomic RNA was isolated from purified virions by digestion with protease K in borate buffer (pH 9) containing 1% SDS and 4 mM EDTA for one hour at 37°C, followed by phenol/chloroform extraction and ethanol precipitation. The RNA was redispersed in water, quantified by absorbance at 260 nm, and analyzed by agarose gel electrophoresis in the presence of methyl mercuric hydroxide.

DNAs complementary to ZYMV RNA were synthesized essentially according to Gubler, U., et al. Gene (1983) 25:263-269, as described in the technical manual for the Riboclone cDNA Synthesis Kit (Promega Corp.). Figure 3 shows an outline of this procedure. The first strand was synthesized using AMV reverse transcriptase and an oligodeoxythymidylate primer. After second strand synthesis EcoRI linkers were added, digested, and the cDNAs were ligated into the EcoRI site of pBluescript (Stratagene, Inc.). The ligation product was then used to transform competent E. coli XL-1 Blue cells (Stratagene) which were then plated in the presence of lac inducer (IPTG) and substrate (X-gal) for color selection of recombinants. Plasmid DNA was isolated from colorless clones by the alkaline lysis miniprep method (Molecular Cloning. A Laboratory Manual, 2d Ed., J. Sambrook, E. Fritsch, and T. Maniatis, eds., Cold Spring Harbor Press, New York, 1989) and insert sizes were

20  
25  
30  
35

SUBSTITUTE SHEET

-28-

estimated after digestion with EcoRI by agarose gel electrophoresis in the presence of ethidium bromide.

PZR1, the largest cDNA clone obtained from the first experiment, had a 2.3 kb insert, the ends of which were sequenced using the Sanger dideoxy chain-terminating method as described in the product literature for the Sequenase 2 sequencing kit (United States Biologicals) with the M13 universal and reverse primers encoded on either end of the multiple cloning site in pBluescript. The orientation of the insert relative to the viral genome and the multiple cloning site was indicated by the appearance of the polyadenylate tract from the 3' end of the genome in the sequence from the reverse primer. The remainder of the clone was sequenced stepwise, 200-300 nucleotides at a time, in both directions from synthetic oligodeoxynucleotide primers complementary to the distal ends of each of the successive sequencing runs. Sequence data were processed and analyzed on a DEC VAX 11/750 minicomputer.

The second round of cDNA cloning was accomplished in the same manner as the first except that a synthetic oligodeoxynucleotide complementary to the 5' end of pZR1 was used as primer and the cDNAs were ligated directly, without linkers, into the EcoRV site of pBluescript (Figure 3). From this cloning two clones were obtained, pZB11 and pZB60, which had inserts of 2.3 kb and 3.8 kb, respectively. For the sequencing of pZB11, nested deletions were prepared from each end of the insert according to Henikoff, S., Gene (1984) 28:357ff as described in the product literature for the Erase-a-base System kit (Promega Corp.). For each direction, approximately twenty-four clones containing deletions spanning the entire length of the insert were sequenced simultaneously from the M13 universal or

**SUBSTITUTE SHEET**

-29-

reverse primers. Any gaps left by failure of the sequences of adjacent time points to overlap were filled in using synthetic oligodeoxynucleotide primers made from the sequence near the 5' end of the gap.

5 Clone pZB60 was sequenced in both directions from nested deletions in the same manner as for pZB11 except that a unique NcoI site within pZB11 was used as the starting point for deletions in the 5' direction and only those clones with deletions mapping between the 5'  
10 end of pZB11 and the 5' end of pZB60 were sequenced.

A third round of cDNA cloning was conducted as described above for the preparation of pZB11 and pZB60 except that a synthetic oligomer complementary to the 5' end of pZB60 was used as a primer. From this round,  
15 pZF18, having an insert of 3.7 kb was obtained and sequenced as described above for pZB11 and pZB60.

The 5' end of the viral RNA sequence was determined by reverse transcription of purified viral RNA using a synthetic oligonucleotide primer complementary to  
20 nucleotides 76-99 at the 5' end of pZF18. The Sanger dideoxynucleotide chain-terminating method was used essentially as described in the Promega Gem Seq manual (Promega Corp.).

The continuous open reading frame of the viral  
25 genome was identified with the aid of a computer as described above. The coding sequences of the functional ZYMV gene products were identified by amino acid sequence homology to those of other potyviruses (Allison, R., et al., Virology (1986) 154:9-20; Domier, L.L., et al.,  
30 Nucleic Acids Res (1986) 14:5417-5430; Robaglia, C., et al., J Gen Virol (1989) 70:935-947; Maiss, B., et al., J Gen Virol (1989) 70:513-524). The identity of the coat protein gene was further confirmed by subcloning the  
35 presumptive coding sequence into a modified version of

**SUBSTITUTE SHEET**

-30-

pBluescript from which the gene could be expressed in vitro. In vitro translation of the gene produced a product of the expected size which reacted specifically with antiserum raised against purified ZYMV coat protein when analyzed by Western blotting.

Figure 4 shows the nucleotide sequence of the ZYMV genome as determined above along with the deduced amino acid sequence. The nucleotide sequence is numbered from the 5' terminus. The 5' non-coding region extends from nucleotide 1 to nucleotide 139. Nucleotides 140-142 initiate the polyprotein coding sequence with a methionine codon in a consensus translation initiation context (Joshi, C.P., Nucleic Acids Res (1987) 15:6643-6653). By homology with the potyviral polyprotein sequences cited above, the cleavage site between the aphid transmission helper component (HC) and the 46 kDa protein is believed to occur between the glycine at codon 766 (nucleotides 2435-2437) and the glycine at codon 767 (nucleotides 2438-2440). The cleavage site between the 46 kDa protein and the cytoplasmic inclusion protein (CI) is believed to occur between the glutamine at codon 1164 (nucleotides 3629-3631) and the glycine at codon 1165 (nucleotides 3632-3634). The cleavage site between CI and VPg/protease (VPg and protease are probably not separated in ZYMV) is believed to occur between the glutamine at codon 1798 (nucleotides 5531-5533) and the serine at codon 1799 (nucleotides 5534-5536). The cleavage site between VPg/protease and RNA replicase (Rep) is believed to occur between the glutamine at codon 2284 (nucleotides 6989-6991) and the serine at codon 2285 (nucleotides 6992-6994). The cleavage site between the RNA replicase and the coat protein (CP) is believed to occur between the glutamine at codon 2801 (nucleotides 8540-8542) and the serine at codon 2802 (nucleotides

**SUBSTITUTE SHEET**



-31-

8543-8545). Termination of the polyprotein coincides with termination of the coat protein and is believed to occur at the stop codon (nucleotides 9380-9382) following the glutamine at codon 3080. The 3' non-coding sequence then extends from nucleotide 9383 to nucleotide 9593 before terminating in a polyadenylate sequence of variable length. cDNA clone pZR1 contained approximately 80 adenosines at its 3' terminus.

10 pZPro5

A restriction fragment of 1666 base pairs (bp) extending between the PvuII and SspI sites of ZYMV cDNA clone pZB11 (described above) was isolated by agarose gel electrophoresis and ligated into the SmaI site of plasmid pTZ18U (Sambrook, J., et al., Molecular Cloning, Cold Spring Harbor Laboratory, 1989; Mead D.A., et al., Protein Engineering (1986) 1:67). This restriction fragment comprises a portion of the coding sequence of the ZYMV polyprotein which includes part of the cytoplasmic inclusion protein (CI), the 6 kDa protein, the 49 kDa protease, and a portion of the N1b protein. Insertion of this fragment into the SmaI site of pTZ18U places the reading frame encoding these proteins in phase with that of the expressible lacZ $\alpha$  gene of pTZ18U such that expression of this gene from the lac promoter is expected to produce a fusion protein comprised of a small portion of the lacZ $\alpha$  peptide fused to the amino terminus of the ZYMV polyprotein fragment. This construct was denoted pZPro5 and its structure was confirmed by dideoxynucleotide sequencing (Sanger, F., et al., Proc. Natl. Acad. Sci. USA (1977) 74:5463-5467).

35 pZPro6 and pZPro7

**SUBSTITUTE SHEET**

-32-

The 2280 bp SalI restriction fragment from ZYMV cDNA clone pZR1 (described above), which comprises a portion of the ZYMV genome including part of the NIb protein, CP, the 3' non-coding sequence, and a portion of the polyadenylate sequence, was inserted into the SalI site of pZPro5 to create pZPro6. Dideoxynucleotide sequencing of pZPro6 confirmed that the NIb/CP-encoding reading frame of the inserted fragment was in phase with the open reading frame of pZPro5 such that expression of this construct from the lac promoter is expected to produce a single polypeptide of approximately 140 kDa.

In a further refinement of pZPro6, the 1244 bp MluI-NaeI fragment was removed and the 705 bp MluI-EcoRV fragment from pZR1 was inserted in its place, removing most of the ZYMV 3' non-coding and polyadenylate sequences, which include several unwanted restriction sites. This construct was denoted pZPro7. *E. coli* strain DH5 $\alpha$  was transformed with pZPro6 and pZPro7, and transformed clones were identified and isolated by selection for ampicillin resistance.

Expression of the lacZ $\alpha$ -ZYMV polyprotein gene in pZPro6 and pZPro7 was monitored by immunoblotting of SDS/PAGE-resolved proteins from these cells using polyclonal antisera raised in rabbits against denatured ZYMV coat protein (Burnette, W.N., Anal. Biochem. (1981) 112:195). Results are shown in Figure 5B. Extract from cells harboring either pZPro6 or pZPro7 contained a single immunoreactive band which co-migrated with the major species of mature ZYMV coat protein at approximately 31 kDa. Since the CP-containing primary translation product of 140 kDa was not detected in these extracts, the exclusive appearance of mature CP implies correct and efficient processing of the polyprotein by the ZYMV 49 kDa protease.

**SUBSTITUTE SHEET**

-33-

To rule out the possibility that mature CP might have been produced either by an endogenous *E. coli* protease, or by fortuitous initiation of translation near the amino terminus of mature CP, a variant of pZPro7 was also analyzed. This variant, denoted placZ $\alpha$ -CP, was made by deleting the sequence encoding the protease and most of the NIB protein from pZPro7, leaving part of the NIB protein and CP in phase with the lacZ $\alpha$  peptide in a 46 kDa open reading frame (see Figure 5A). Extracts from cells harboring this construct contained a single immunoreactive band which migrated with an apparent MW of 46 kDa (Figure 5B, lane 2). The apparent absence in these cells of a species co-migrating with mature CP in the absence of the ZYMV 49 kDa protease indicates that the activity of the latter is indeed responsible for the occurrence of mature CP in cells harboring pZPro6 and pZPro7.

#### Example 2

Construction and analysis of genes which confer a negative phenotype on *E. coli* cells by virtue of the activity of the ZYMV 49 kDa protease according to the scheme described above for Protease Inhibitor Selection System I.

This example describes the construction of expressible genes encoding polyproteins which contain the ZYMV 49 kDa protease and the *E. coli* ribosomal protein S12. The ability of these gene constructs to confer sensitivity to the antibiotic streptomycin on several streptomycin-resistant *E. coli* strains by virtue of correct and efficient processing of the polyprotein by the ZYMV 49 kDa protease is demonstrated.

Streptomycin lethality in *E. coli* has been ascribed to its ability to interfere with protein

**SUBSTITUTE SHEET**

-34-

synthesis by binding to the S12 subunit of the 30S component of the ribosome (Gorini, L., in Ribosomes, M. Nomura, A. Tissieres, P. Lengyel, eds., Cold Spring Harbor Laboratory, 1974, pp. 791-803). Streptomycin-resistant mutants have been isolated which express altered forms of S12 which retain the ability to participate in the assembly of ribosomes which can function in the presence of streptomycin. Wildtype S12 has been shown to confer a dominant streptomycin-sensitive phenotype on merodiploids which express both wildtype and streptomycin-resistant forms of S12. The highly sequestered position of S12 in the ribosome suggests that S12-containing polyproteins should be too encumbered to participate in the assembly of functional ribosomes. Thus, the expression of such polyproteins in streptomycin-resistant hosts should be unable to confer streptomycin sensitivity unless mature S12 can be proteolytically freed from the polyprotein.

#### 20 pZPro9

The CP-encoding sequences in pZPro7 were precisely replaced with the coding sequence for S12 to create pZPro9. This was accomplished as follows. The sequence bounded by the BglII site in N1b and the P1' position of the N1b/CP cleavage site in pZPro7 (Schecter, I., and A. Berger, Biochem. Biophys. Res. Commun. (1967) 27:157) was amplified by polymerase chain reaction (PCR, Saiki, R.K., et al., Science (1988) 239:487). The S12 coding sequence from the second amino acid to the end was amplified by PCR from plasmid pNO1523 (Dean, D., Gene (1981) 15:99-102). The 3' primer contained an MluI site following the stop codon. Following cleavage of the first PCR product by BglII and the second by MluI both were simultaneously ligated into pZPro7 from which the

**SUBSTITUTE SHEET**

-35-

BglII-MluI fragment had been removed. Following transformation and plasmid DNA purification, the structure of pZPro9 (see Figure 6) was confirmed by dideoxynucleotide sequencing.

5 pZPro9 was then transformed into streptomycin-resistant *E. coli* strains MC1009, HB101, and N100 (American Type Culture Collection Catalogue of Bacteria and Phages, 1989). After two rounds of single colony isolation in the presence of ampicillin (amp), single  
10 colonies of each transformant were grown in Luria-Bertani medium (LB) containing 50 µg/ml amp to mid-log phase and plated on solid LB containing 100 µg/ml amp, 100 µg/ml streptomycin (strep), or both.

Consistently, fewer than one in  $10^5$  amp-resistant colony-forming units (cfu) of each transformant  
15 was observed to grow in the presence of both amp and strep, while the same hosts harboring pZPro7 plated with similar efficiencies on amp alone or amp and strep (see Figure 6). Thus, by virtue of having S12 in place of CP, pZPro9 is able to confer strep sensitivity on strep-resistant hosts, while its CP-containing parent, pZPro7,  
20 is not. The pZPro9 transformants were fully sensitive to as little as 3 µg strep/ml while the parent strains were fully resistant to up to 350 µg/ml. Also, the pZPro9  
25 transformants plated equally well on strep alone or amp alone, indicating that pZPro9 is quickly lost in the absence of amp selection and that there is no discernible tendency to replace the strep-resistant gene in the host chromosome with the S12 gene by homologous recombination.

30 To confirm that cleavage of the polyprotein by the ZYMV 49 kDa protease to liberate mature S12 is required for strep sensitivity, the N1b/CP cleavage site was removed from pZPro9 to create pZPro12, which should  
35 produce a polyprotein from which S12 cannot be freed by

**SUBSTITUTE SHEET**

-36-

the protease. This was accomplished by cleaving pZPro9 with EcoRV and HpaI, which removed most of the NIB protein including the NIB/CP cleavage site, and replacing it with the fragment produced by EcoRV alone, which  
5 restored most of the NIB protein down to within 12 amino acids of the NIB/CP cleavage site (see Figure 6). The structure of pZPro12 was confirmed by dideoxynucleotide sequencing. Transformation with pZPro12 has no discernible effect on the ability of strep-resistant *E. coli*  
10 *coli* strains to grow vigorously in the presence of up to 350 µg/ml streptomycin. Thus, the strep-sensitive phenotype produced by pZPro9 is completely dependent on the presence of a substrate cleavage site at which the protease can cleave functional S12 from the polyprotein.

15 The expression of the pZPro9 polyprotein, like most large eucaryotic proteins, places a considerable burden on growing *E. coli* cells. In an attempt to reduce this burden, the pZPro9 was streamlined by removing the EcoRV fragment described above, which contains most of  
20 the NIB protein exclusive of the cleavage sites at either end. This construct, denoted pZPro10, encodes a polyprotein of about 83 kDa, of which 49 kDa is the protease and about 14 kDa is S12 (see Figure 6). Upon transformation with pZPro10, strep-resistant *E. coli*  
25 strains displayed a strep-sensitive phenotype identical to that of the pZPro9 transformants. In addition, pZPro10 transformants grew considerably more vigorously than pZPro9 transformants, indicating a significant reduction in the metabolic burden on the host cells.  
30 Thus, removal of most of the NIB protein had no discernible effect on the efficiency of removal of functional S12 from the polyprotein by the protease.

35 However, pZPro10-expressing cells still grew poorly compared to the untransformed host. Recent work

SUBSTITUTE SHEET

-37-

with another viral protease suggests that this is probably due, at least in part, to fortuitous activity of the protease on host proteins (Baum, E.Z., et al., Proc. Natl. Acad. Sci. USA (1990) 87:5573-5577). Inhibitors of the protease should at least partly restore normal growth. As this growth differential is relatively easy to score, it is possible to use the toxicity of the protease as the negative phenotype, and to select inhibitors from peptide libraries, or to confirm selected inhibitors on the basis of their ability to restore rapid growth.

Once the protease removes itself from the polyprotein of pZPro10, the 14 kDa S12 is left in a 21.5 kDa precursor until freed by the protease. To confirm that neither this precursor nor the polyprotein itself is able to confer strep sensitivity, the N1b/CP cleavage site was removed from pZPro10 in the same manner that the EcoRV-HpaI restriction fragment was removed from pZPro9 to create pZPro12. This new construct, pZPro11, shown in Figure 6, had no discernible effect on the level of strep resistance shown by strep-resistant E. coli strains. Thus, again, the presence of a substrate cleavage site adjacent to S12 is required for strep sensitivity, implying that the ZYMV 49 kDa protease is specifically responsible for generating functional S12.

Thus, systems for identifying and selecting protease inhibitors from peptide libraries have been disclosed. Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

35

-38-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

(i) APPLICANT: BALINT, ROBERT

(ii) TITLE OF INVENTION: A METHOD FOR THE IDENTIFICATION  
OF

10

PROTEASE INHIBITORS

(iii) NUMBER OF SEQUENCES: 1

(iv) CORRESPONDENCE ADDRESS:

15

(A) ADDRESSEE: Irell &amp; Manella

(B) STREET: 545 Middlefield Road, Suite 200

(C) CITY: Menlo Park

(D) STATE: CA

(E) COUNTRY: USA

20

(F) ZIP: 94025

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

25

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/727,837

30

(B) FILING DATE: 08-JUL-1991

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

35

(A) NAME: ROBINS, ROBERTA L.

SUBSTITUTE SHEET



-39-

(B) REGISTRATION NUMBER: 33,208

(C) REFERENCE/DOCKET NUMBER: 7115-0045

(ix) TELECOMMUNICATION INFORMATION:

- 5 (A) TELEPHONE: (415) 327-7250  
(B) TELEFAX: (415) 327-2951  
(C) TELEX: 706141

10 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9593 base pairs  
(B) TYPE: nucleic acid  
15 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

25

30

35

**SUBSTITUTE SHEET**

-40-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	AAAATTGAAA CAAATCACAA AGACTACAAG AATCAACGAT CAAGCAAACG AATTTTGGAA	60
5	CGTATTTTACA AACAAAGCAAT CTAAACTCT TACAGTATTA AGAAATTCTC CAATCACTTC	120
	GTTTACTTCA GACATAACAA TGGCCTCCAT CATGATTGGT TCAATCTCTG TACCCATTGC	180
	AAAGACTGAG CAGTGTGCAA ACACTCAAGT AAGTAATCGG GCTAATATAG TGGCACCTGG	240
10	CCACATGGCA ACATGCCCCAT TGCCACTGAA AACGCACATG TATTACAGGC ATGAGTCCAA	300
	GAAGTTGATG CAATCAAACA AGAGCATTGA CATTCTGAAC AACTTCTTCA GCACTGAOGA	360
	GATGAAGTTT AGGCTCACTC GAAACGAGAT GAGCAAGCTG AAAAAGGGTC CGAGCGGGAG	420
15	GATAGTCCTC CGCAAGCCGA GTAAGCAGCG GGTTTTCGCT CGTATCGAGC AGGATGAGGC	480
	AGCACGCAAG GAAGAGGCTG TTTTCCTCGA AGGAAATTAT GACGATTCCA TCACAAATCT	540
	AGCACGTGTT CTTCACCTG AAGTGACTCA CAACGTTGAT GTGAGCTTGC GATCACCGTT	600
20	TTACAAGCGC ACATACAAGA AGGAAAGGAA GAAAGTGGCG CAAAAGCAAA TTGTGCAAGC	660
	ACCACTTAAT AGCTTGTGCA CACGTGTTCT TAAATTGCA CGCAATAAAA ATATCCCTGT	720
	TGAGATGATT GGCAACAAGA AGGCGAGACA TACACTCACC TTCAAGAGGT TTAGGGGATG	780
25	TTTTGTTGGA AAGGTGTCAG TTGCGCATGA AGAAGGACGA ATGCGGCACA CTGAGATGTC	840
	GTATGAGCAG TTTAAATGGC TTCTTAAAGC CATTGTCTAG GTCACCCATA CAGAGCGAAT	900
	TCGTGAGGAA GATATTAAAC CAGGTTGTAG TGGGTGGGTG TTGGGCACTA ATCATACATT	960
30	GACTAAAAGA TATTCAAGAT TGCCACATT GGTGATTGCA GGTAGAGACC ACGATGGGAT	1020
	TGTGAACCGG CTGGAACAGG TGTATTTTA TAGCGAAGTT GACCACTATT CGTCGCAACC	1080
35	GGAAGTTCAG TTCTTCCAAG GATGGCGACG AATGTTTGAT AAGTTTAGGC CTAGCCCAGA	1140

SUBSTITUTE SHEET

-41-

	TCATGTGTGC AAAGTTGACC ACAACAACGA GGAATGTGGT GAGTTAGCAG CAATCTTTTG	1200
	TCAGGCTCTA TTCCCAGTAG TGAACTATC GTGCCAACA TGCAGAGAAA AGCTTAGTAG	1260
5	AGTTAGCTTT GAGGAATTCA AAGATTCTTT GAACGCAAC TTTATTATCC ACAAGGATGA	1320
	ATGGGGTAGT TTCAAGGAAG GCTCTCAATA CGATAATATT TTCAAATTAA TCAAAGTGGC	1380
	AACACAGGCA ACTCAGAATC TCAAGCTCTC ATCTGAAGTT ATGAAATTAG TTCAGA <sup>?</sup> ACCA	1440
10	CACAAGCACT CACATGAAGC AAATACAAGA CATCAATAAG GCGCTCATGA AAGGTTTATT	1500
	GCTTGCGCAA GACGAATTGG ACTTAGCTTT GAAACAGCTT CTTGAAATGA CTCAGTGGTT	1560
	TAAGAACCAC ATGCACCTGA CTGGTGAGGA GGCATTGAAG ATGTTCAGAA ATAAGCGTTC	1620
15	TAGCAAGGCC ATGATAAATC CTAGCCTTCT ATGTGGCAAC CAATTGGACA AAAATGGAAA	1680
	TTTTGTTTGG GGAGAAAGAG GATACCATTG CAAGCGATTA TTCAAGAACT TCTTCCAAGA	1740
	AGTAATACCA AGCGAAGGAT ATACGAAGTA CGTAGTCCGA AACTTTCCAA ATGGTACTCG	1800
20	TAAGTTGGCC ATAGGCTCAT TGATTGTACC ACTTAATTG GATAGGGCAC GCACTGCACT	1860
	ACTTGGAGAG AGTATTGAGA AGAAGCCACT CACATCAGCG TGTGTCTCCC AACAGAATGG	1920
	AAATTATATA CACTCATGCT GCTGTGTAAC GATGGATGAT GGAACCCCGA TGTACTCCGA	1980
25	GCTTAAGAGC CCGACGAAGA GGCATCTAGT TATAGGAGCT TCTAGTGATC CAAAGTACAT	2040
	TGATCTGCCA GCATCTGAGG CAGAACGCAT GTATATAGCA AAGGAAGGTT ATTGCTATCT	2100
	CAGTATTTTC CTCGCAATGC TTGTAAATGT TAATGAGAAC GAAGCAAAGG ATTCACCAA	2160
30	AATGATTGCT GATGTTTTGA TCCCATGCT TGGGCAGTGG CCTTCATTGA TGGATGTTGC	2220
	AACTGCAGCA TATATTCTAG GTGTATTCCA TCCTGAAACG CGATGCGCTG AATTACCCAG	2280
	GATCCTTGTT GACCACGCTA CACAAACCAT GCATGTCATT GATTCTTATG GATCACTAAC	2340
35	TGTTGGGTAT CACGTGCTCA AGGCTGGAAC TGCAATCAT TTAATTCAAT TTGCCTCAAA	2400

SUBSTITUTE SHEET

-42-

5  
10  
15  
20  
25  
30  
35

TGATCTGCAA AGCGAGATGA AACATTACAG AGTTGGTGGG ACACCAACAC AGCGCATTAA 2460  
ACTCGAGGAG CAGCTGATTA AAGGAATTTT CAAACCAAAA CTTATGATGC AGCTCCTGCA 2520  
TGATGACCCA TACATATTAT TACTTGGCAT GATTTCACCC ACCATTCTTG TACATATGTA 2580  
TAGGATGCGT CATTTTGAGC GGGGTATTGA GATATGGATT AAGAGGGATC ATGAAATCGG 2640  
AAAGATTTTC GTCATATTAG AGCAGCTCAC ACGCAAGGT GCTCTGGCAG AAGTTCTTGT 2700  
GGATCAACTT AACTTGATAA GTGAAGCTTC ACCACATTTA CTTGAAATTA TGAAGGGTTG 2760  
TCAAGATAAT CAGAGGGCAT ACGTACCTGC GCTGGATTTC CTAACGATAC AAGTGGAGCG 2820  
TGAGTTTTC AATAAAGAAC TCAAAACCAA TGGCTATCCA GATTTCAGC AAACGCTCTT 2880  
CGATATGAGG GAAAAAATGT ATGCAAAGCA GCTGCACAAT TCATGGCAAG AGCTAAGCTT 2940  
GCTGGAAAAA TCCTGTGTAA CCGTGGGATT GAAGCAATTC TCGATTTTTC CGGAAAGAAA 3000  
TTTAATCCAG CGAGCAAAAG AAGGAAAGCG CGCATCTTCG CTACAATTG TTCACGAGTG 3060  
TTTTATCAG ACCCGAGTAC ATGCGAAGAG CATTGCGGAT GCAGGCGTGC GTAAACTAAA 3120  
TGAGGCTCTC GTCGGAACCT GTAAATCTT TTTCTCTTGT GGTTCAAAA TTTTGGCGG 3180  
ATGCTATAGC GACATAATAT ACCTTGTAAG CGTGTGTTTG GTTTTCTCCT TGGTGCTACA 3240  
AATGTCCAAT ACTGTGCGCA GTATGATAGC AGCGACAAGG GAAGAAAAAG AGAGAGCGAT 3300  
GGCAAAATAA GCTGATGAAA ATGAAAGGAC GTTAATGCAT ATGTACCACA TTTTCAGCAA 3360  
GAAACAGGAT GATGCGCCCA TATACAATGA CTTTCTTGAA CATGTGCGTA ATGTGAGACC 3420  
AGATCTTGAG GAACTCTCT TGTACATGGC TGGCGTAGAA GTTGTTTCAA CACAGGCTAA 3480  
GTCAGCGGTT CAGATTCAAT TCGAGAAAAT TATAGCTGTG TTGGCGCTGC TTACCATGTG 3540  
CTTTGACGCC GAAAGAAGCG ATGCCATTTT CAAGATTTTG ACAAACCTCA AAACAGTTTT 3600  
TGGTACGGTT GGAGAAACGG TCCGACTTCA AGGGCTTGAA GACATTGAAA GCTTGGAGGA 3660

SUBSTITUTE SHEET

-43-

	CGATAAAAGA CTCACAATTG ATTTTGATAT TAACACGAAC GAGGCTCAAT CGTCAACAAC	3720
	ATTTGATGTC CATTTTGATG ACTGGTGGAA TCGGCAACTA CAGCAAAATC GCACAGTTCC	3780
5	ACATTACAGG ACCACAGGCA AATTCCTTGA ATTTACCAGA AATACTGCAG CTTTGTGGC	3840
	CAATGAAATA GCATCATCAA GTGAGGGAGA GTTCTTAGTT AGAGGAGCAG TAGGTTCTGG	3900
	AAAATCAACG AGCTTACCTG CACATCTTGC CAAGAAGGGT AAGGTGTTAC TACTCGAACC	3960
10	TACACGCCCT TTGGCGGAGA ATGTTAGTAG ACAGTTAGCA GGTGATCCTT TCTTCAAAA	4020
	CGTTACACTC AGAATGAGAG GGTAAAGTTG TTTTGGTTCA AGCAATATTA CAGTGATGAC	4080
	GAGTGGATTT GCTTTTCACT ACTATGTTAA CAATCCACAT CAATTGATGG AATTGACTT	4140
15	TGTCTCATA GACGAGTGCC ATGTCACAGA CAGTCCGACC ATAGCTTTCA ATTGTGCACT	4200
	TAAAGAGTAC AACTTTGCTG GCAAATTGAT TAAAGTGTCT GCAACGCGC CAGGGAGAGA	4260
	GTGCGATTTC GATACGCAAT TCGCGGTGAA AGTCAAAACA GAGGACCATC TTTCAATCCA	4320
20	TGCATTGTTT GCGGCACAGA AGACTGTTT AAATGCTGAC ATGGTTGAGC ATGGTAATAA	4380
	CATACTTGTG TATGTTGCAA GTTACAACGA AGTGGACATG CTCTTAAGT TACTCACTGA	4440
	GCGCCAATTT TCAGTTACAA AGGTAGATGG GCGAACAATG CAGCTTGGAA AACTACCAT	4500
25	TGAAACGCAT GGAAGTAGCC AAAAGCCCCA TTTCATAGTA GCTACAAACA TCATCGAGAA	4560
	TGGAGTGACC TTGGATGTTG AGTGTGTTGT TGATTTTGGG CTAAAAGTGG TCGCAGAACT	4620
	GGACAGCGAA AATCGGTGTG TCGGCTACAA TAAGAAATCA GTTAGTTATG GAGAGAGGAT	4680
30	TCAGCGACTA GGAAGAGTGG GGAGATCTAA GCCTGGAAGT GCATTGCGTA TAGGGCACAC	4740
	AGAAAAAGGC ATCGAAACCA TTCCTGAATT CATTGCCACA GAAGCAGCAG CCTTATCATT	4800
	TGCATATGGG CTCCAGTCA CCACACATGG AGTTCCACA AATATACTTG GAAAGTGCAC	4860
35	AGTTAAACAG ATGAAATGTG CTTTGAACCT TGAGCTAACT CCTTTCTTCA CCACTCATT	4920

SUBSTITUTE SHEET

-44-

5  
10  
15  
20  
25  
30  
35

AATCCGTCAT GATGGTAGTA TGCATCCACT AATACACGAA GAATTGAAGC AGTTCAAACT 4980  
CAGGGATTCA GAAATGGTGC TCAACAAGGT TGCATTACCT CATCAATTTG TGAGCCAATG 5040  
GATGGATCAA AGTGAGTATG AACGCATTGG AGTGCACGTT CAATGCCATG AGAGCACACG 5100  
CATACCTTTT TACACAAATG GAATACCTGA TAAAGTCTAT GAGAGAATTT GGAAGTGCAT 5160  
ACAAGAAAAC AAGAACGATG CGGTTTTTGG TAAGCTTTCA AGTGCTTGTT CAACTAAGGT 5220  
TAGTTATACA CTTAGCACTG ATCCAGCAGC ATTACCCAGA ACTATTGCAA TCATCGATCA 5280  
CCTGCTTGCC GAGGAAATGA TGAAGCGGAA TCACTTCGAC ACTATCAGCT CAGCTGTAAC 5340  
GGGCTATTCA TTTTCCCTTG CTGGAATTGC TGATTCTTTC AGGAAGAGAT ACATGCGCGA 5400  
TTACACAGCG CACAACATTG CAATTCTCCA ACAAGCACGT GCCCAGCTGC TTGAATTTAA 5460  
TAGTAAGAAT GTGAACATTA ACAATCTGTC CGATTTAGAA GGAATTGGAG TCATTAAGTC 5520  
GGTGGTGTTC CAAAGTAAGC AAGAGGTCAG CAGTTTCCTC GGACTTCGGG GTAAATGGGA 5580  
TGGAAAGAAA TTTGCGAATG ATGTGATATT GCGGATTATG ACACTCTTAG GAGGTGGGTG 5640  
GTTTATGTGG GAATACTTCA CGAAAAAGAT CAATGAACCC GTGCGCGTTG AAAGCAAGAA 5700  
ACGTCGATCT CAAAAATTGA AATTCAGGGA TGGGTACGAT AGAAAAGTTG GACGTGAGAT 5760  
TTTTGGTGAT GATGATACAA TTGGGCGCAC TTTGCGCGAA GCTTACACGA AGAGAGGAAA 5820  
GGTCAAAGGA AACAAACA CAAAAGGAAT GGGACGGAAA ACTCGCAATT TTGTGCATTT 5880  
ATATGGTGTG GAGCCTGAGA ATTACAGTTT TATCAGATTT GTGGACCCTC TCACTGGCCA 5940  
TACATTGGAC GAAAGCACCC ATACAGACAT ATCGTTAGTG CAGGAGGAGT TTGGAAGTAT 6000  
TAGAGAGAAA TTTCTGGAGA ATGATTTGAT CTCGAGGCAG TCTATTATCA ACAAACCCGG 6060  
CATTGAGCA TATTTTATGG GCAAGGGCAC TGAAGAAGCA CTCAAAGTTG ACTTGACTCC 6120  
TCATGTACCA TTGCTTCTGT GCAGAAACAC CAATGCTATT GCGGGATACC CAGAGAGAGA 6180

SUBSTITUTE SHEET

-45-

5  
10  
15  
20  
25  
30  
35

AAATGAGTTG AGACAACTG GCACACCAGT CAAGGTTTCT TTTAAAGACG TGCCAGAGAA 6240  
AAACGAACAT GTCGAGTTGG AGAGCAAATC TATCTACAA GGAGTGC GCG ATTACAATGG 6300  
CATCTCAACA ATCGTTTGT C AATTAACGAA CGATTCTGAT GGCCTCAAGG AGACCATGTA 6360  
TGGTATTGGC TATGGGCCAA TAATCATCAC TAATGGACAC CTCTTCAGGA AAAACAATGG 6420  
CACACTTCTA GTCAGGTCTT GGCATGGTGA ATTCATTGTT AAAAATACCA CAACGCTCAA 6480  
AGTGCAATTC ATAGAAGGGA AGGATGTCGT GTTAGTGCGC ATGCCAAAGG ACTTTCGGCC 6540  
GTTTAAAGC AACGCTTCTT TTAGGGCACC AAAACGCGAG GAACGACGAT GCTTG GTTGG 6600  
GACAACTTT CAAGAAAAGA GTCTTCGCTC CACTGTTTCG GAATCTTCCA TGACAATACC 6660  
TGAAGGAACT GGCTCATATT GGATACATTG GATTTCGACC AACGAAGGGG ATTGCGGATT 6720  
GCCCATGGTT TCAACAACGG ATGGCAAGAT AATTGGAGTT CATGGTTTGG CTCCACAGT 6780  
CTCATCTAAG AATTATTTTG TCCCATTAC TGATGATTTT ATAGCCACGC ATTTGAGCAA 6840  
ACTTGATGAC CTCACATGGA CTCAGCATTG GCTATGGCAA CCTAGCAAAA TTGCGTGGGG 6900  
AACGCTCAAC TTAGTTGATG AACCAACCAGG GCCCGAATT CGTATCTCAA ATCTAGTCAA 6960  
GGATTTATTC ACTTCTGGTG TTGAAACACA GAGCAAGCGA GAAAGATGGG TCTACGAAAG 7020  
CTGTGAAGGG AACCTTCGGG CTGTTGGAAC TGCACAATCA GCGTTAGTCA CCAAACATGT 7080  
TGTGAAAGGC AAGTGTCTT TCTTGAAGA ATATTTACAA ACACACGCAG AAGCGAGCGC 7140  
CTATTTGAGA CCCCTAATGG GAGAGTACCA GCCGAGCAAG TTGAACAAAG AAGCCTTTAA 7200  
AAAGGATTTT TTTAAATACA ATAAACCCGT CACTGTTAAC CAACTGGATC ATGATAAATT 7260  
TTTGGGAGCA GTGGATGGGG TTATACGTAT GATGTGTGAT TTTGAGTTCA ACGAATGTG 7320  
ATTCATTACA GATCCCGAGG AAATTACAA CTCTTTGAAC ATGAAAGCAG CAATTGGAGC 7380  
CCAGTATAGA CGAAAGAAGA AAGAGTATTT TGAGGGGCTA GATGATTTTG ATCGAGAGCG 7440

SUBSTITUTE SHEET

-46-

ACTTTTATTC CAAAGTTGTG AAAGGTTGTT CAATGGCTAC AAAGGTCTGT GGAATGGATC 7500  
TTTAAAGGCC GAGCTCAGGC CGCTTGAGAA AGTCAGGGCT AACAAAACAC GAACCTTTAC 7560  
AGCAGCGCCA ATTGATACAT TGCTTGGAGC TAAAGTTTGT GTGGATGATT TCAACAATGA 7620  
5 GTTCTACAGG AAAAACCTCA AGTGTCCATG GACGGTCGGC ATGACAAAAT TTTATGGTGG 7680  
TTGGGATAAA TTGATGAGAT CATTACCTGA TGGTTGGTTG TATTGTCATG CTGATGGATC 7740  
ACAGTTCGAT AGTTCGTTAA CCCCAGCCTT ACTGAACGCA GTGCTCATAA TCAGGTCATT 7800  
10 TTATATGGAG GATTGGTGGG TCGGCCAAGA GATGCTTGAA AATCTTTATG CCGAGATTGT 7860  
GTACACTCCA ATTCTTGCTC CTGATGGAAC AATTTTCAAG AAATTTAGAG GTAACAACAG 7920  
TGGGCAACCC TCAACAGTGG TGGATAACAC ACTAATGGTT GTGATCTCTA TTTACTATGC 7980  
15 GTGCATGAAA TTTGGTTGGA ACTGCGAGGA GATTGAGAAT AACTTGTCT TCTTTGCAAA 8040  
TGGAGATGAT CTGATACTTG CAGTCAAAGA TGAGGATAGC GGCTTACTTG ATAACATGTC 8100  
ATCCTCTTTT TCGGAACCTG GACTGAATTA TGATTTTCA GAACGTACGC ATAAAAGAGA 8160  
20 AGATCTTTGG TTCATGTCCC ACCAAGCAAT GCTAGTTGAT GGAATGTACA CTCCAAACT 8220  
CGAGAAAGAG AGAATTGTTT CAATTCTAGA GTGGGATAGA AGCAAAGAAA TTATGCACCG 8280  
AACAGAGGCT ATTTGCGCTG CGATGATTGA GGCATGGGGG CACACCGAGC TCTTGCAAGA 8340  
25 AATCAGAAAG TTTTACCTAT GGTTCGTTGA AAAAGAAGAG GTCCGAGAAT TGGCAGCCCT 8400  
CGGAAAAGCT CCATACATAG CTGAGACAGC ACTTCGTAAG TTATACACTG ACAAGGGAGC 8460  
AGATACAAGT GAACTGGCAC GCTACCTACA AGCCCTCCAT CAAGATATCT TCTTTGAGCA 8520  
30 AGGAGACACT GTGATGCTCC AATCAGGCAC TCAGCCAACT GTGGCAGATG CTGGAGCTAC 8580  
AAAGAAAGAT AAAGAAGATG ACAAAGGGAA AAACAAGGAC GTTACAGGCT CCGGCTCAGG 8640  
35 TGAGAAAACA GTAGCAGCTG TCACGAAGGA CAAGGATGTG AATGCTGGTT CTCATGGGAA 8700

SUBSTITUTE SHEET



-47-

	AATTGTGCCG CGTCTTTTGA AGATCACAAA GAAAATGTCA TTGCCACGCG TGAAGGAAA	8760
	TGTGATACTC GATATTGATC ATTTGCTGGA ATATAAACCG GATCAAATTG AGTTATATAA	8820
5	CACACGAGCG TCTCATCAGC AGTTGCGCTC TTGGTTCAAC CAGGTTAAGA CGGAATATGA	8880
	TTTGAACGAG CAACAGATCG GAGTTGTAAT GAATGGTTTC ATGGTTTGGT GCATTGAGAA	8940
	TGGCACTTCA CCCGACATTA ATGGAGTGTG GGTATGATG GACGGAAATG AGCAAGTTGA	9000
10	GTATCCCTTG AAACCAATAG TTGAAAATGC AAAGCCAAAC CTGCGGCAAA TAATGCATCA	9060
	TTTTTCAGAT GCAGCGGAGG CATATATAGA GATGAGAAAT GCAGAGGCAC CATACTGCC	9120
	GAGGTATGGT TTGCTTCGAA ACCTACGGGA TAGGAGTTTA GCACGATATG CTTTTGATT	9180
15	CTATGAAGTC AATTCTAAAA CTCCTGAAAG AGCCCGCGAA GCTGTTGCGC AGATGAAAGC	9240
	AGCAGCTCTT AGCAATGTTT CTTCAAGGTT GTTTGGCCTT GATGGAAATG TTGCCACCAC	9300
	TAGCGAAGAC ACTGAACGGC ACACTGCACG TGATGTTAAT AGAAACATGC ACACCTTACT	9360
20	AGGTGTGAAT ACAATGCAGT AAAGGGTAGG CCGCCTACCT AGGTTATTGT TTCGCTGCCG	9420
	ACGTAATTCT AATATTTACC GCTTTATTG ATATCTTTAG ATTTCCAGAG TGGGCCTCCC	9480
	ACCTTTAAAG CGTAAAGTTT ATGTTAGTTG TCCAGGAGTG CCGTAGTCCT TTCGGAAGCT	9540
25	TTAGTGTGAG CCTCTCAGG ATAAGCTCGA GATTAGACTC CGTTTGCAAG CCT	9593
30		
35		

SUBSTITUTE SHEET

-48-

Claims

1. A method for detecting a protease inhibitor, said method comprising:
- 5 (a) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on said host cells dependent on the activity of said
- 10 protease;
- (b) providing a pool of nucleic acid constructs wherein at least one of said constructs in said pool comprises a nucleic acid sequence encoding an inhibitor of said protease;
- 15 (c) transforming said host cells of (a) with said nucleic acid constructs of (b); and
- (d) growing said transformed host cells of (c) under conditions that distinguish cells with said selectable phenotype, thereby detecting the presence of
- 20 said protease inhibitor.
2. The method of claim 1 wherein said host cells are bacterial cells.
- 25 3. The method of claim 2 wherein said selectable phenotype is the ability of said bacterial cells to grow in the presence of a given antibiotic.
- 30 4. The method of claim 3 wherein said second nucleic acid sequence comprises a nucleic acid sequence encoding E. coli ribosomal protein S12 and said antibiotic is streptomycin.

35

-49-

5. The method of claim 2 wherein said first nucleic acid sequence comprises a nucleic acid sequence encoding ZYMV 49 kDa protease.

5 6. The method of claim 2 wherein said selectable phenotype is the ability of said transformed host cells to grow in the presence of a given carbon source.

10 7. The method of claim 1 wherein said second nucleic acid sequence comprises a nucleic acid sequence encoding a protein which is inactivated by said protease.

15 8. The method of claim 1 wherein said second nucleic acid sequence comprises a nucleic acid sequence encoding a protein which is activated by said protease.

9. A DNA construct comprising:

20 (a) a first DNA coding sequence for a protein capable of conferring a selectable phenotype on a host cell transformed therewith, said selectable phenotype dependent on the activity of a protease; and

(b) control sequences that are operably linked to said first and second coding sequences whereby said  
25 coding sequences can be transcribed and translated in a host cell, and at least one of said control sequences is heterologous to at least one of said coding sequences.

10. The DNA construct of claim 9 further  
30 comprising a second DNA coding sequence for said protease.

35 11. The DNA construct of claim 10 wherein said protease is ZYMV 49 kDa protease.

**SUBSTITUTE SHEET**

-50-

12. The DNA construct of claim 9 wherein said first DNA coding sequence codes for E. coli ribosomal protein S12 and said selectable phenotype is streptomycin resistance.

5

13. The DNA construct of claim 10 wherein said first DNA coding sequence codes for E. coli ribosomal protein S12 and said selectable phenotype is streptomycin resistance.

10

14. The DNA construct of claim 11 wherein said first DNA coding sequence codes for E. coli ribosomal protein S12 and said selectable phenotype is streptomycin resistance.

15

15. A DNA construct comprising:

(a) a DNA coding sequence for a protein capable of inhibiting the action of a given protease, said protein identified by the method of claim 1; and

20

(b) control sequences that are operably linked to said coding sequence whereby said coding sequence can be transcribed and translated in a host cell, and at least one of said control sequences is heterologous to at least said coding sequence.

25

16. The DNA construct of claim 15 wherein said protease is ZYMV 49 kDa protease.

17. A host cell stably transformed with a DNA construct according to claim 9.

30

18. A host cell stably transformed with a DNA construct according to claim 10.

35

-51-

19. The host cell of claim 18 further transformed with a DNA construct comprising:

(a) a DNA coding sequence for a protein capable of inhibiting the action of a given protease, said protein identified by a method comprising

5 (i) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on  
10 said host cells dependent on the activity of said protease;

(ii) providing a pool of nucleic acid constructs wherein at least one of said constructs in said pool comprises a nucleic acid sequence encoding an  
15 inhibitor of said protease;

(iii) transforming said host cells of (i) with said nucleic acid constructs of (ii); and

(iv) growing said transformed host cells of (iii) under conditions that distinguish cells with  
20 said selectable phenotype, thereby detecting the presence of said protease inhibitor; and

(b) control sequences that are operably linked to said coding sequence whereby said coding sequence can be transcribed and translated in a host cell, and at  
25 least one of said control sequences is heterologous to at least said coding sequence.

20. A host cell stably transformed with a DNA construct according to claim 11.

30

21. The host cell of claim 20 further transformed with a DNA construct comprising:

35

-52-

(a) a DNA coding sequence for a protein capable of inhibiting the action of a given protease, said protein identified by a method comprising

5 (i) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on said host cells dependent on the activity of said protease;

10 (ii) providing a pool of nucleic acid constructs wherein at least one of said constructs in said pool comprises a nucleic acid sequence encoding an inhibitor of said protease;

15 (iii) transforming said host cells of (i) with said nucleic acid constructs of (ii); and

(iv) growing said transformed host cells of (iii) under conditions that distinguish cells with said selectable phenotype, thereby detecting the presence of said protease inhibitor; and

20 (b) control sequences that are operably linked to said coding sequence whereby said coding sequence can be transcribed and translated in a host cell, and at least one of said control sequences is heterologous to at least said coding sequence.

25

22. A host cell stably transformed with a DNA construct according to claim 12.

30 23. A host cell stably transformed with a DNA construct according to claim 13.

24. The host cell of claim 23 further transformed with a DNA construct comprising:

35

**SUBSTITUTE SHEET**

-53-

(a) a DNA coding sequence for a protein capable of inhibiting the action of a given protease, said protein identified by a method comprising

- 5 (i) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on said host cells dependent on the activity of said protease;
- 10 (ii) providing a pool of nucleic acid constructs wherein at least one of said constructs in said pool comprises a nucleic acid sequence encoding an inhibitor of said protease;
- 15 (iii) transforming said host cells of (i) with said nucleic acid constructs of (ii); and
- (iv) growing said transformed host cells of (iii) under conditions that distinguish cells with said selectable phenotype, thereby detecting the presence of said protease inhibitor; and
- 20 (b) control sequences that are operably linked to said coding sequence whereby said coding sequence can be transcribed and translated in a host cell, and at least one of said control sequences is heterologous to at least said coding sequence.

25

25. A host cell stably transformed with a DNA construct according to claim 14.

26. The host cell of claim 25 further
- 30 transformed with a DNA construct comprising:

(a) a DNA coding sequence for a protein capable of inhibiting the action of ZYMV 49 kDa protease, said protein identified by a method comprising

35

-54-

(i) providing a population of host cells expressing a first nucleic acid sequence encoding a protease and a second nucleic acid sequence encoding a protein capable of conferring a selectable phenotype on said host cells dependent on the activity of said protease;

(ii) providing a pool of nucleic acid constructs wherein at least one of said constructs in said pool comprises a nucleic acid sequence encoding an inhibitor of said protease;

(iii) transforming said host cells of (i) with said nucleic acid constructs of (ii); and

(iv) growing said transformed host cells of (iii) under conditions that distinguish cells with said selectable phenotype, thereby detecting the presence of said protease inhibitor; and

(b) control sequences that are operably linked to said coding sequence whereby said coding sequence can be transcribed and translated in a host cell, and at least one of said control sequences is heterologous to at least said coding sequence.

25

30

35

SUBSTITUTE SHEET



1 / 5

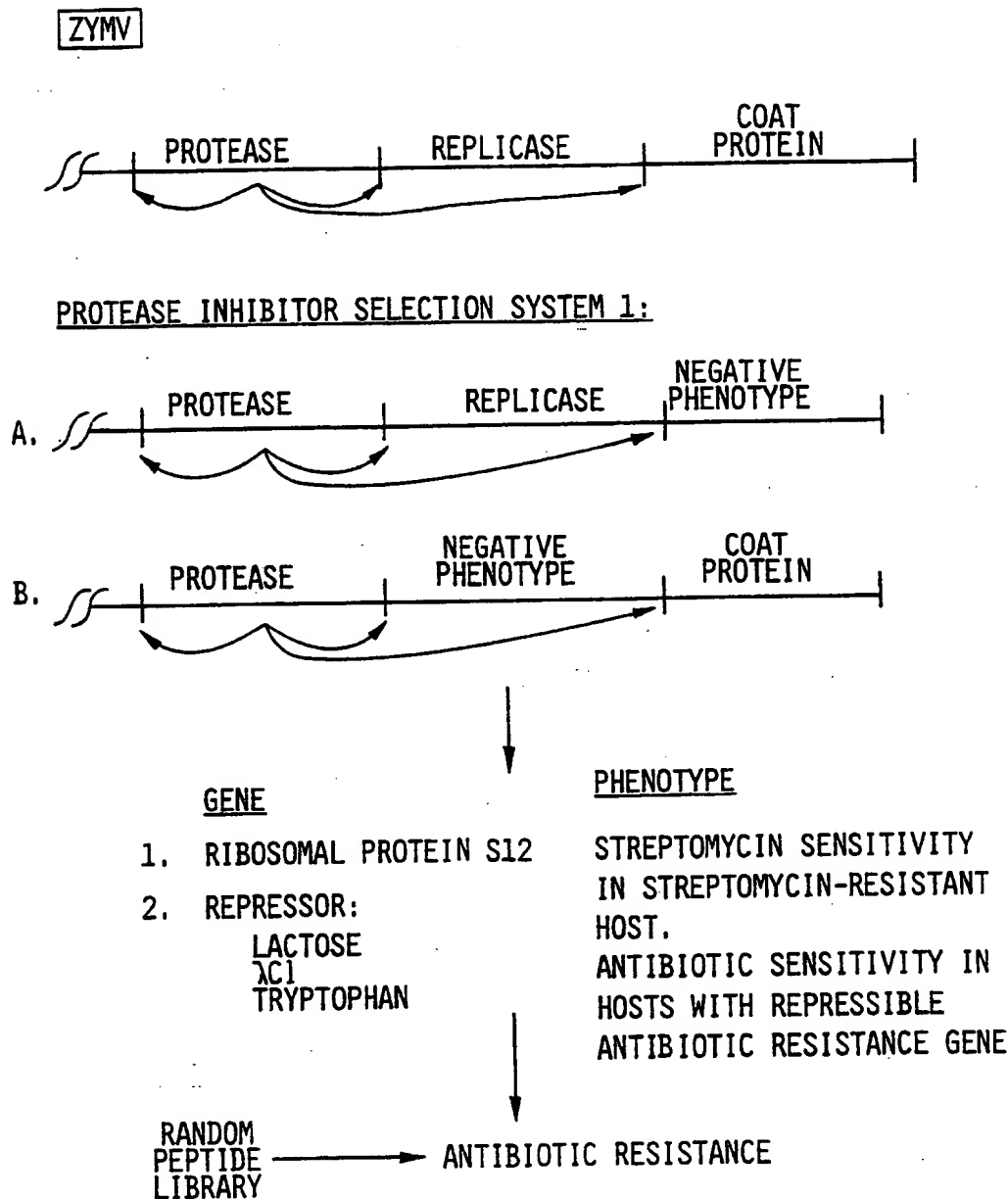
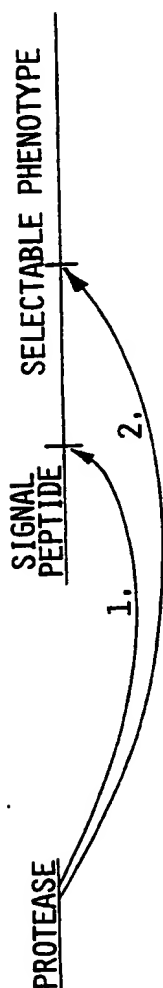


FIG. 1

2 / 5



1. SECRETED OR MEMBRANE PROTEINS WITH CLEAVAGE SITE AFTER SIGNAL PEPTIDE.

- A. ANTIBIOTIC RESISTANCE-  
 B-LACTAMASE (AMPICILLIN)  
 TETRACYCLINE  
 NEOMYCIN PHOSPHOTRANSFERASE (KANAMYCIN)  
 B. CARBON SOURCE UTILIZATION-LAC PERMEASE (LACTOSE)  
 Malt (MALTULOSE)

2. CLEAVAGE SITE INSERTED INTO SURFACE DOMAIN OF ANTIBIOTIC RESISTANCE OR  
 CARBON SOURCE UTILIZATION GENE.

FIG. 2

3 / 5

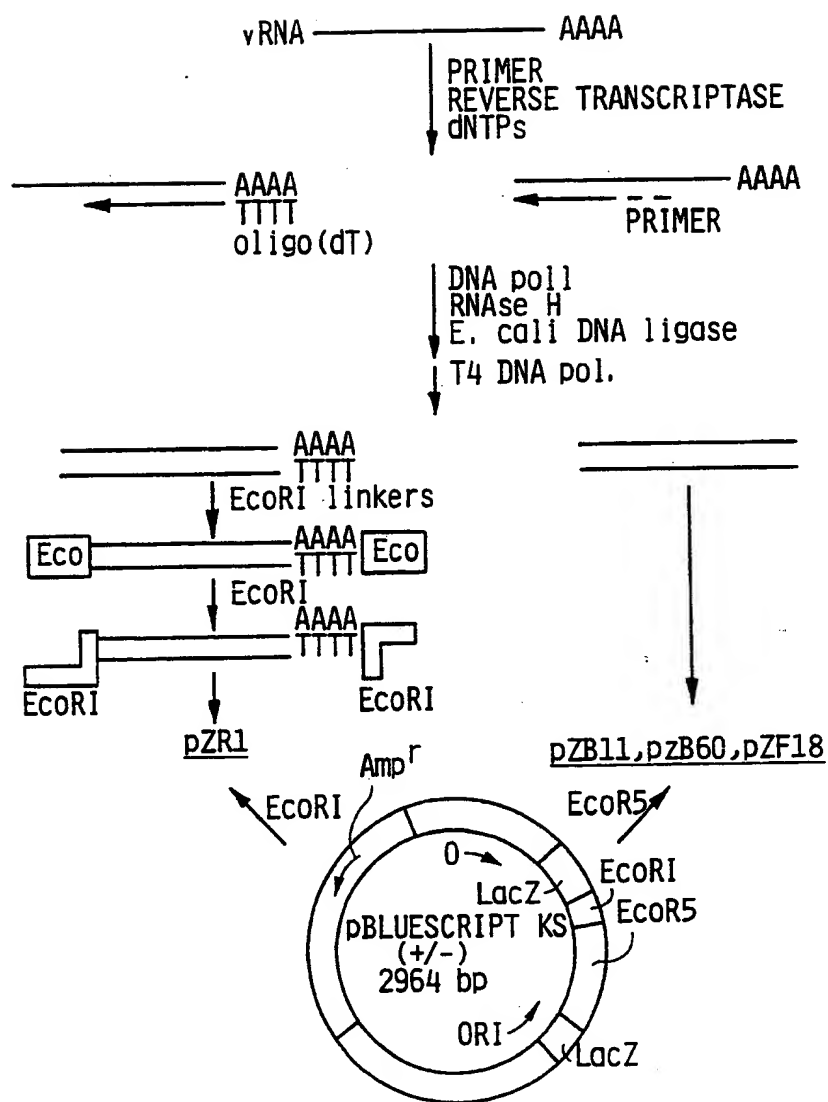


FIG. 3

SUBSTITUTE SHEET

4 / 5

FIG. 5A

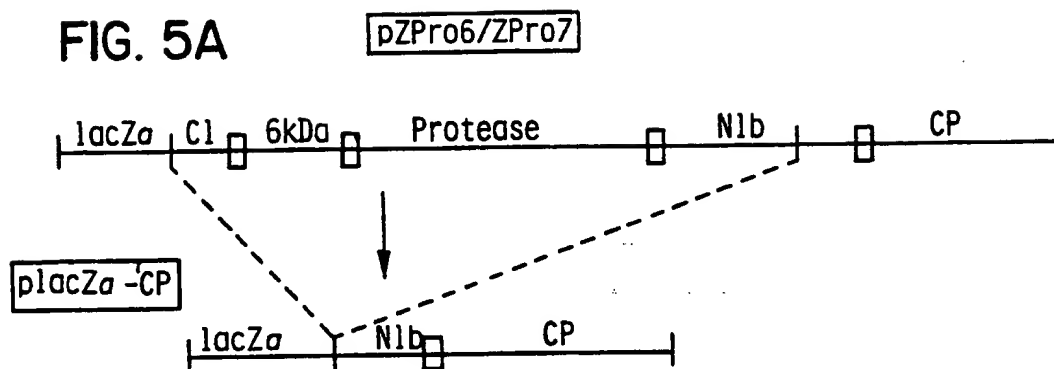
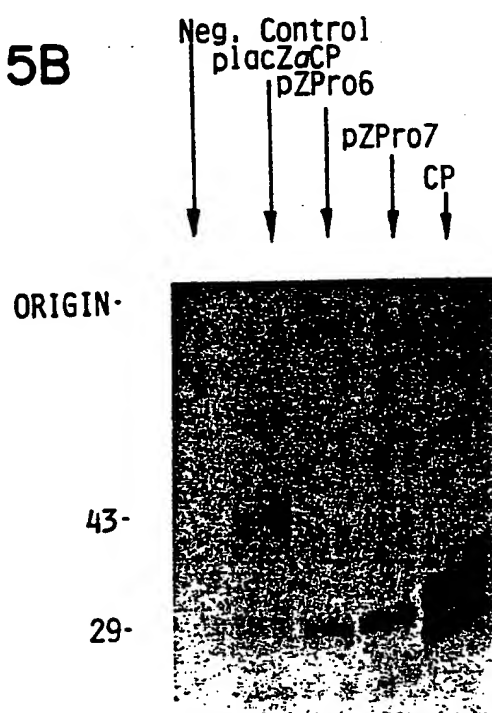
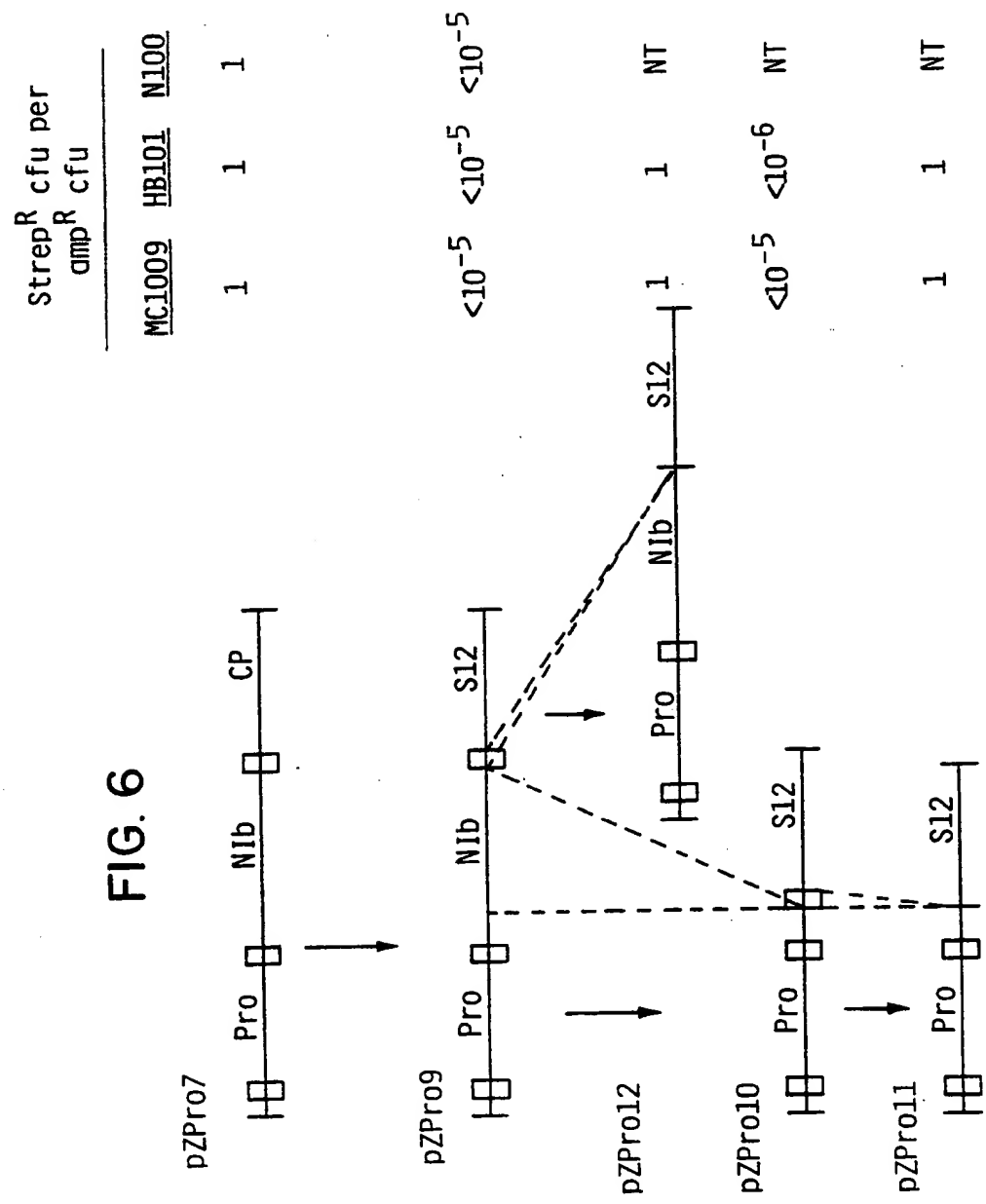


FIG. 5B



SUBSTITUTE SHEET

FIG. 6



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 92/05745

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12Q 1.02; C12N 15/15, 1.21

US CL : 435/7.37, 320.1, 252.33; 536/27

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.37, 320.1, 252.33; 536/27

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, BIOSIS, WPI

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the National Academy of Sciences, Volume 87, issued December 1990, E. Z. Baum et al., "β-Galactosidase containing a human immunodeficiency virus protease cleavage site is cleaved and inactivated by human immunodeficiency virus protease", pages 10023-10027, entire document.	1-26
Y	EP, A, 0,321,973 (Sommergruber et al.) 28 June 1989, entire document.	1-26
Y	Journal of General Virology, Volume 71, issued 1990, R. Grumet et al., "cDNA cloning and sequence analysis of the 3'-terminal region of zucchini yellow mosaic virus RNA", see especially Figure 1.	1-26
Y	Gene, Volume 15, issued 1981, D. Dean, "A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids", see entire document.	1-26



Further documents are listed in the continuation of Box C.



See patent family annex.

•

Special categories of cited documents:

\*A\*

document defining the general state of the art which is not considered to be part of particular relevance

\*E\*

earlier document published on or after the international filing date

\*L\*

document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\*

document referring to an oral disclosure, use, exhibition or other means

\*P\*

document published prior to the international filing date but later than the priority date claimed

\*T\*

later document published after the international filing date or priority date and not in conflict with the application but cited to underscore the principle or theory underlying the invention

\*X\*

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*Z\*

document member of the same patent family

Date of the actual completion of the international search

25 August 1992

Date of mailing of the international search report

31 AUG 1992

Name and mailing address of the ISA/  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

JAMES KETTER

Telephone No. (703) 308-0196